

Persistence and uncertainty in the academic career

Alexander M. Petersen^{a,1}, Massimo Riccaboni^{b,c,d}, H. Eugene Stanley^{a,1}, and Fabio Pammolli^{a,b,c,e,1}

^aLaboratory for the Analysis of Complex Economic Systems, Institutions Markets Technologies (IMT) Lucca Institute for Advanced Studies, 55100 Lucca, Italy; ^bLaboratory of Innovation Management and Economics, IMT Lucca Institute for Advanced Studies, 55100 Lucca, Italy; ^cCrisis Lab, IMT Lucca Institute for Advanced Studies, 55100 Lucca, Italy; ^dDepartment of Managerial Economics, Strategy and Innovation, Katholieke Universiteit Leuven, 3000 Leuven, Belgium; and ^eCenter for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215

Contributed by H. Eugene Stanley, December 28, 2011 (sent for review December 14, 2011)

Understanding how institutional changes within academia may affect the overall potential of science requires a better quantitative representation of how careers evolve over time. Because knowledge spillovers, cumulative advantage, competition, and collaboration are distinctive features of the academic profession, both the employment relationship and the procedures for assigning recognition and allocating funding should be designed to account for these factors. We study the annual production $n_i(t)$ of a given scientist i by analyzing longitudinal career data for 200 leading scientists and 100 assistant professors from the physics community. Our empirical analysis of individual productivity dynamics shows that (i) there are increasing returns for the top individuals within the competitive cohort, and that (ii) the distribution of production growth is a leptokurtic “tent-shaped” distribution that is remarkably symmetric. Our methodology is general, and we speculate that similar features appear in other disciplines where academic publication is essential and collaboration is a key feature. We introduce a model of proportional growth which reproduces these two observations, and additionally accounts for the significantly right-skewed distributions of career longevity and achievement in science. Using this theoretical model, we show that short-term contracts can amplify the effects of competition and uncertainty making careers more vulnerable to early termination, not necessarily due to lack of individual talent and persistence, but because of random negative production shocks. We show that fluctuations in scientific production are quantitatively related to a scientist's collaboration radius and team efficiency.

career trajectory | labor market | science of science | tenure | computational sociology

Institutional change could alter the relationship between science and scientists as well as the longstanding patronage system in academia (1, 2). Some recent shifts in academia include the changing business structure of research universities (3), shifts in the labor supply demand balance (4), a bottleneck in the number of tenure track positions (5), and a related policy shift away from long-term contracts (3, 6). Along these lines, significant factors for consideration are the increasing range in research team size (7), the economic organization required to fund and review collaborative research projects, and the evolving definition of the role of the academic research professor (3).

The role of individual performance metrics in career appraisal, in domains as diverse as sports (8, 9), finance (10, 11), and academia, is increasing in this data rich age. In the case of academia, as the typical size of scientific collaborations increases (7), the allocation of funding and the association of recognition at the varying scales of science [individual \leftrightarrow group \leftrightarrow institution (12)] has become more complex. Indeed, scientific achievement is becoming increasingly linked to online visibility in a considerable reputation tournament (13).

Here we seek to identify (i) quantitative patterns in the scientific career trajectory towards a better understanding of career dynamics and achievement (14–20), and (ii) how scientific production responds to policies concerning contract length. Using rich productivity data available at the level of single individuals, we analyze longitudinal career data keeping in mind the roles

of spillovers, group size, and career sustainability. Although our empirical analysis is limited to careers in physics, our approach is general. We speculate that similar features describe other disciplines where academic publication is a primary indicator and collaboration is a key feature.

Specifically, we analyze production data for 300 physicists $i = 1 \dots 300$ who are distributed into 3 groups: (i) Group A corresponds to the 100 most cited physicists with average h -index $\langle h \rangle = 61 \pm 21$, (ii) Group B corresponds to 100 additional highly cited physicists with $\langle h \rangle = 44 \pm 15$, and (iii) Group C corresponds to 100 assistant professors in 50 US physics departments with $\langle h \rangle = 15 \pm 7$. We define the annual production $n_i(t)$ as the number of papers published by scientist i in year t of his/her career. We focus on academic careers from the physics community to approximately control for significant cross-disciplinary production variations. Using the same set of scientists, a companion study has analyzed the rank-ordered citation distribution of each scientist with a focus on the statistical regularities underlying publication impact (17). We provide further description of the data and present a parallel analysis of 21,156 sports careers in *SI Appendix*.

We begin this paper with empirical analysis of longitudinal career data. Our empirical evidences serve as statistical benchmarks used in the final section where we develop a stochastic proportional growth model. In particular, our model shows that a short-term appraisal system can result in a significant number of “sudden” early deaths due to unavoidable negative production shocks. This result is consistent with a Matthew Effect model (16) and recent academic career survival analysis (21), which demonstrate how young careers can be stymied by the difficulty in overcoming early achievement barriers. Altogether, our results indicate that short-term contracts may increase the strength of the “rich-get-richer” mechanism in science (22, 23) and may hinder the upward mobility of young scientists.

Results

Scientific Production and the Career Trajectory. The academic career depends on many factors, such as cumulative advantage (16, 19, 22, 23), the “sacred spark,” (24, 25), and other complex aspects of knowledge transfer manifest in our techno-social world (26). To exemplify this complexity, a recent case study on the impact trajectories of Nobel prize winners shows that “scientific career shocks” marked by the publication of an individual's “magnum opus” work(s) can trigger future recognition and reward, resembling the cascading dynamics of earthquakes (27).

We model the career trajectory as a sequence of scientific outputs which arrive at the variable rate $n_i(t)$. Because the reputation of a scientist is typically a cumulative representation of his/her

Author contributions: A.M.P., M.R., H.E.S., and F.P. designed research; A.M.P., M.R., H.E.S., and F.P. performed research; A.M.P., M.R., H.E.S., and F.P. analyzed data; and A.M.P., M.R., H.E.S., and F.P. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence may be addressed. E-mail: hes@bu.edu, Alexander M. Petersen alexander.petersen@imtlucca.it, or Fabio Pammolli f.pammolli@imtlucca.it.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1121429109/-DCSupplemental.

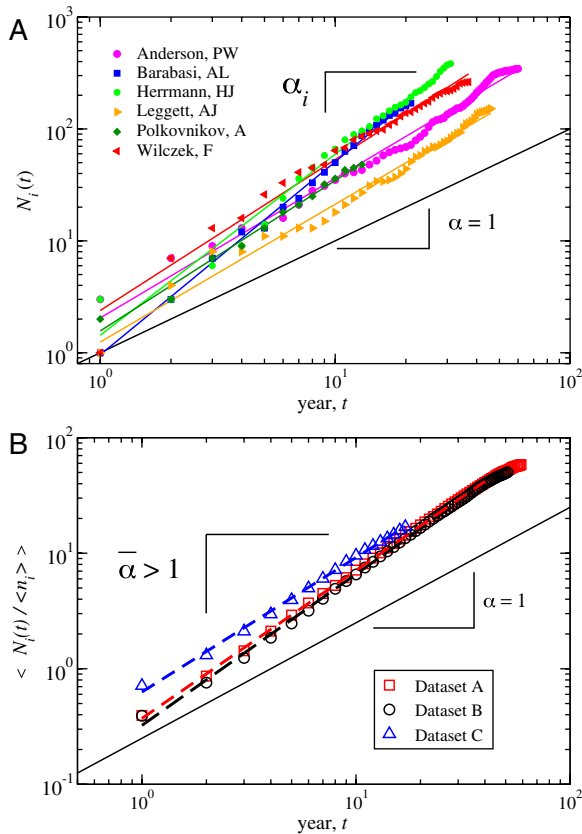


Fig. 1. Persistent accelerating career growth. (A) The career trajectory $N_i(t) \sim t^{\alpha_i}$ of six stellar careers from varying age cohorts. The α_i value characterizes the career persistence, where careers with $\alpha > 1$ are accelerating. α_i values calculated using OLS regression in alphabetical order are: $\alpha = 1.25 \pm 0.02$, $\alpha = 1.72 \pm 0.02$, $\alpha = 1.62 \pm 0.04$, $\alpha = 1.23 \pm 0.02$, $\alpha = 1.34 \pm 0.05$, $\alpha = 1.35 \pm 0.04$. (B) Defined in Eq. 1, the average career trajectory $\langle N'(t) \rangle$ calculated from 100 individual $N_i(t)$ in each dataset demonstrates robust accelerating career growth within each cohort. We use the normalized career trajectory $N'_i(t)$ in order to aggregate $N_i(t)$ with varying publication rates $\langle n_i \rangle$. As a result, the aggregate scaling exponent $\bar{\alpha}$ quantifies the acceleration of the typical career over time, independent of $\langle n_i \rangle$. For the scientific careers, we calculate $\bar{\alpha}$ values: 1.28 ± 0.01 [A], 1.31 ± 0.01 [B], and 1.15 ± 0.02 [C]. These values are all significantly greater than unity, $\bar{\alpha} > 1$, indicating that cumulative advantage in science is closely related to knowledge and production spillovers. We calculate $\bar{\alpha}$ using OLS regression and plot the corresponding best-fit lines (dashed) for each dataset.

contributions, we consider the cumulative production $N_i(t) \equiv \sum_{t'=1}^t n_i(t')$ as a proxy for career achievement. Fig. 1A shows the cumulative production $N_i(t)$ of six notable careers which display a temporal scaling relation $N_i(t) \approx A_i t^{\alpha_i}$ where α_i is a scaling exponent that quantifies the career trajectory dynamics. The average and standard deviation of the α_i values calculated for each dataset are $\langle \alpha_i \rangle = 1.42 \pm 0.29$ [A], 1.44 ± 0.26 [B], and 1.30 ± 0.31 [C]. We justify this two-parameter model in the *SI Appendix* text using scaling methods and data collapse.

There are also numerous cases of $N_i(t)$ which do not exhibit such regularity (see *SI Appendix: Fig. S1*), but instead display marked nonstationarity and nonlinearity arising from significant exogenous career shocks. Positive shocks, possibly corresponding to just a single discovery, can spur significant productivity and reputation growth (24, 27). Negative shocks, such as in the case of scientific fraud, can end the career rather suddenly. We also acknowledge that the end of the career is a difficult phase to analyze, because such an event can occur quite abruptly, and so our analysis is mainly concerned with the growth phase and not the termination phase.

In order to analyze the average properties of $N_i(t)$ for all 300 scientists in our sample, we define the normalized trajectory $N'_i(t) \equiv N_i(t)/\langle n_i \rangle$. The quantity $\langle n_i \rangle$ is the average annual production of author i , with $N'_i(L_i) = L_i$ by construction (L_i corresponds to the career length of individual i). Fig. 1B shows the characteristic production trajectory obtained by averaging together the 100 $N'_i(t)$ belonging to each dataset,

$$\langle N'(t) \rangle \equiv \frac{\langle N_i(t) \rangle}{\langle n_i \rangle} \equiv \frac{1}{100} \sum_{i=1}^{100} \frac{N_i(t)}{\langle n_i \rangle}. \quad [1]$$

The standard deviation $\sigma(N'(t))$ shown in *SI Appendix: Fig. S2B* begins to decrease after roughly 20 y for dataset [A] and [B] scientists. Over this horizon, the stochastic arrival of career shocks can significantly alter the career trajectory (20, 24, 27, 28).

Each $N'_i(t)$ exhibits robust scaling corresponding to the scaling law $\langle N'(t) \rangle \sim t^{\bar{\alpha}}$. This regularity reflects the abundance of careers with $\alpha_i > 1$ corresponding to accelerated career growth. This acceleration is consistent with increasing returns arising from knowledge and production spillovers.

Fluctuations in Scientific Output over the Academic Career. Individuals are constantly entering and exiting the professional market, with birth and death rates depending on complex economic and institutional factors. Due to competition, decisions and performance at the early stages of the career can have long lasting consequences (16, 29). To better understand career uncertainty portrayed by the common saying “publish or perish” (30), we analyze the outcome fluctuation

$$r_i(t) \equiv n_i(t) - n_i(t - \Delta t) \quad [2]$$

of career i in year t over the time interval $\Delta t = 1$ y. Fig. 2A and B show the unconditional probability density function (pdf) of r values which are leptokurtic but remarkably symmetric, illustrating the endogenous frequencies of positive and negative output growth. Output fluctuations arise naturally from the lulls and bursts in both the mental and physical capabilities of humans (31, 32). Moreover, the statistical regularities in the annual production change distribution indicate a striking resemblance to the growth rate distribution of countries, firms, and universities (33, 34).

To better account for individual growth factors, we next define the normalized production change

$$r'_i(t) \equiv [r_i(t) - \langle r_i \rangle] / \sigma_i(r) \quad [3]$$

which is measured in units of the fluctuation scale $\sigma_i(r)$ unique to each career. We measure the average $\langle r_i \rangle$ and the standard deviation $\sigma_i(r)$ of each career using the first L_i available years for each scientist i . $r'_i(t)$ is a better measure for comparing career uncertainty, because individuals have production factors that depend on the type of research, the size of the collaboration team, and the position within the team. Fig. 2C shows that $P(r')$, the pdf of r' measured in units of standard deviation, is well approximated by a Gaussian distribution with unit variance. The data collapse of each $P(r')$ onto the predicted Gaussian distribution (solid green curve) indicates that individual output fluctuations are consistent with a proportional growth model. We note that the remaining deviations in the tails for $|r'| \geq 3$ are likely signatures of the exogenous career shocks that are not accounted for by an endogenous proportional growth model.

The ability to collaborate on large projects, both in close working teams and in extreme examples as remote agents [i.e. Wikipedia (35)], is one of the foremost properties of human society. In science, the ability to attract future opportunities is strongly related to production and knowledge spillovers

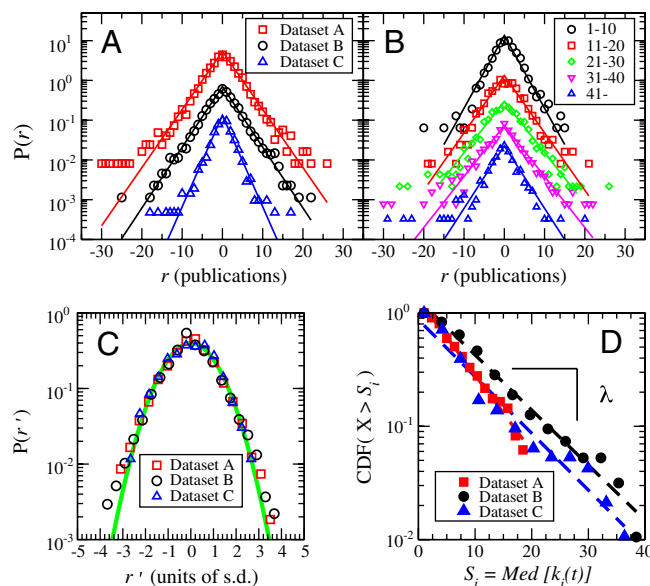


Fig. 2. Empirical evidence for the proportional growth model of career production. (A) pdf of the annual production change r in the number of papers published over a $\Delta t = 1$ y period. In the bulk of each $P(r)$, the growth distribution is approximately double-exponential (Laplace). (B) To test the stability of the distribution over career trajectory subintervals, we separate $r_i(t)$ values into 5 nonoverlapping 10-year periods and verify the stability of the Laplace $P(r)$. For each $P(r)$, we also plot the corresponding Laplace distribution (solid line) with standard deviation σ and mean $\mu \approx 0$ calculated using the maximum likelihood estimator method. To improve graphical clarity, we vertically offset each $P(r)$ by a constant factor. For visual comparison, we also plot a Normal distribution (dashed black curve) with $\sigma \equiv 1$ which instead decays parabolically on the log-linear axes. (C) Accounting for individual production factors by using the normalized production change r' , the resulting pdfs $P(r')$ collapse onto a Gaussian distribution with unit variance. Deviations in the tails likely correspond to extreme “career shocks.” (D) The cumulative distribution $CDF(X \geq S_i)$ is exponential, indicating that the unconditional distributions $P(r)$ in (A) and (B) follow from an exponential mixing of conditional Gaussian distributions $P(r|S_i)$.

(28, 36, 37) that are facilitated by the collaboration network (7, 12, 38–42). Indeed, there is a tipping point in a scientific career that occurs when a scientist’s knowledge investment reaches a critical mass that can sustain production over a long horizon, and when a scientist becomes an attractor (as opposed to a pursuer) of new collaboration/production opportunities. To account for collaboration, we calculate for each author the number $k_i(t)$ of distinct coauthors per year and then define his/her collaboration radius S_i as the median of the set of his/her $k_i(t)$ values, $S_i \equiv \text{Med}[k_i(t)]$. We use the median instead of the average $\langle k_i(t) \rangle$ because extremely large $k_i(t)$ values can occur in specific fields such as high-energy physics and astronomy.

Given the complex scientific coauthorship network, we ask the question: what is the typical number of unique coauthors per year? Fig. 2D shows the cumulative distribution function $CDF(S_i)$ of S_i values for each dataset. The approximately linear form on log-linear axes indicates that S_i is exponentially distributed, $P(S_i) \sim \exp[-\lambda S_i]$. We calculate $\lambda = 0.15 \pm 0.01$ [A], $\lambda = 0.11 \pm 0.01$ [B], and $\lambda = 0.11 \pm 0.01$ [C]. The exponential size distribution has been shown to emerge in complex systems where linear preferential attachment governs the acquisition of new opportunities (43). This result shows that the leptokurtic “tent-shaped” distribution $P(r)$ in Fig. 2 follows from the exponential mixing of heterogeneous conditional Gaussian distributions (44).

The exponential mixture of Gaussians decomposes the unconditional distribution $P(r)$ into a mixture of conditional Gaussian distributions

$$P(r|S_i) = \exp[-r^2/2VS_i^\psi]/\sqrt{2\pi VS_i^\psi}, \quad [4]$$

each with a fluctuation scale $\sigma_i(r)$ depending on S_i by the scaling relation

$$\sigma_i^2(r) \approx V S_i^\psi. \quad [5]$$

Hence, the mixture is parameterized by ψ

$$P_{\psi}(r) = \int_0^{\infty} P(r|S)P(S)dS \approx \sum_{i=1} P_i(r|S_i)P(S_i). \quad [6]$$

The independent case $\psi = 0$ results in a Gaussian $P_\psi(r)$ and the linear case $\psi = 1$ results in a Laplace (double-exponential) $P_\psi(r)$. See [SI Appendix](#) and ref. 44 for further discussion of the ψ dependence of $P_\psi(r)$.

The Size-Variance Relation and Group Efficiency. The values of ψ for scientific and athletic careers follow from the different combination of physical and intellectual inputs that enter the production function for the two distinct professions. Academic knowledge is typically a nonrival good, and so knowledge-intensive professions are characterized by spillovers, both over time and across collaborations (36, 37), consistent with $\alpha_i > 1$ and $\psi > 0$. Interestingly, Azoulay, et al. show evidence for production spillovers in the 5–8% decrease in output by scientists who were close collaborators with a “superstar” scientists who died suddenly (28).

We now formalize the quantitative link between scientific collaboration (38, 39) and career growth given by the size-variance scaling relation in Eq. 5 visualized in the scatter plot in Fig. 3*B*. Using ordinary least squares (OLS) regression of the data on log-log scale, we calculate $\psi/2 \approx 0.40 \pm 0.03$ ($R = 0.77$) for dataset [A], $\psi/2 \approx 0.22 \pm 0.04$ ($R = 0.51$) [B], and $\psi/2 \approx 0.26 \pm 0.05$ ($R = 0.45$) [C]. Interdependent tasks that are characteristic of group collaborations typically involve partially overlapping efforts. Hence, the empirical ψ values are significantly less than the value $\psi = 1$ that one would expect from the sum of S_i independent random variables with approximately equal variance V . Collectively, these empirical evidences serve as coherent motivations for the preferential capture growth model that we propose in the following section.

Alternatively, it is also possible to estimate ψ using the relation between the average annual production $\langle n_i \rangle$ and the collaboration radius S_i . The input-output relation $\langle n_i \rangle \sim S_i^\psi$ quantifies the collaboration efficiency, with $\psi = 0.74 \pm 0.04$ ($R = 0.87$) for dataset [A] and $\psi = 0.25 \pm 0.04$ ($R = 0.37$) for dataset [B]. If the autocorrelation between sequential production values $n_i(t)$ and $n_i(t+1)$ is relatively small, then we expect the scaling exponents calculated for $\langle n_i \rangle$ and $\sigma_i^2(r)$ to be approximately equal. This result follows from considering $r_i(t)$ as the convolution of an underlying production distribution $P_i(n)$ for each scientist that is approximately stable. Interestingly, the larger ψ values calculated for dataset [A] scientists suggests that prestige is related to the increasing returns in the scientific production function (45).

Next we use an alternative method to estimate the annual collaboration efficiency by relating the number of publications $n_i(t)$ in a given year to the number of distinct coauthors $k_i(t)$ over the same year. We use a single-factor production function,

$$n_i(t) \approx q_i[k_i(t)]^{\gamma_i}, \quad [7]$$

to quantify the relation between output and labor inputs with a scaling exponent γ_i . We estimate q_i and γ_i for each author using OLS regression, and define the normalized output measure $Q_i \propto n_i(t)/k_i(t)^{\gamma_i}$ using the best-fit q_i and γ_i values calculated for each scientist i . Fig. 3C shows the efficiency parameter γ

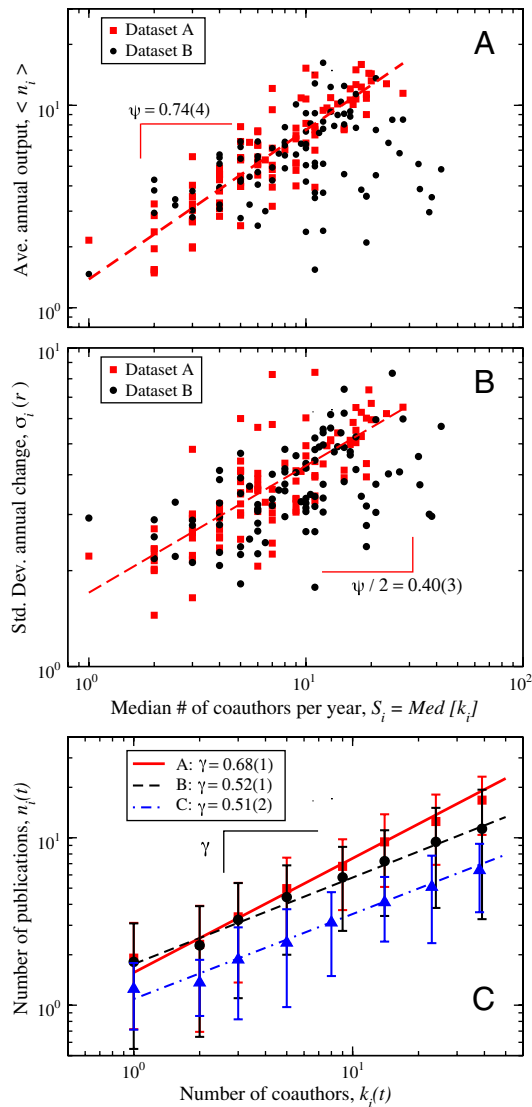


Fig. 3. Quantitative relations between career growth, career risk, and collaboration efficiency. The fluctuations in production reflect the unpredictable horizon of “career shocks” which can affect the ability of a scientists to access new creative opportunities. (A) Relation between average annual production (\bar{n}_i) and collaboration radius $S_i \equiv \text{Med}[k_i]$ shows a decreasing marginal output per collaborator as demonstrated by sublinear $\psi < 1$. Interestingly, dataset [A] scientists have on average a larger output-to-input efficiency. (B) The production fluctuation scale $\sigma_i(r)$ is a quantitative measure for uncertainty in academic careers, with scaling relation $\sigma_i(r) \sim S_i^{\psi/2}$. (C) Management, coordination, and training inefficiencies can result in a $\gamma < 1$ corresponding to a decreasing marginal return with each additional coauthor input. The significantly larger γ value for dataset [A] scientists seems to suggest that managerial abilities related to output efficiency is a common attribute of top scientists.

calculated by aggregating all careers in each dataset, and indicates that this aggregate γ is approximately equal to the average $\langle \gamma_i \rangle$ calculated from the γ_i values in each career dataset: $\gamma = 0.68 \pm 0.01$ [A], $\gamma = 0.52 \pm 0.01$ [B], and $\gamma = 0.51 \pm 0.02$ [C]. Furthermore, the ψ and γ values are approximately equal, which is not surprising, because both scaling exponents are efficiency measures that relate the scaling relation of output $n_i(t)$ per input $k_i(t)$.

A Proportional Growth Model for Scientific Output. We develop a stochastic model as a heuristic tool to better understand the effects of long-term vs. short-term contracts. In this competition

model, opportunities (i.e., new scientific publications) are captured according to a general mechanism whereby the capture rate $\mathcal{P}_i(t)$ depends on the appraisal $w_i(t)$ of an individual’s record of achievement over a prescribed history. We define the appraisal to be an exponentially weighted average over a given individual’s history of production

$$w_i(t) \equiv \sum_{\Delta t=1}^{t-1} n_i(t - \Delta t) e^{-c\Delta t}, \quad [8]$$

which is characterized by the appraisal horizon $1/c$. We use the value $c = 0$ to represent a long-term appraisal (tenure) system and a value $c \gg 1$ to represent a short-term appraisal system. Each agent $i = 1 \dots I$ simultaneously attracts new opportunities at a rate

$$\mathcal{P}_i(t) = \frac{w_i(t)^\pi}{\sum_{i=1}^I w_i(t)^\pi} \quad [9]$$

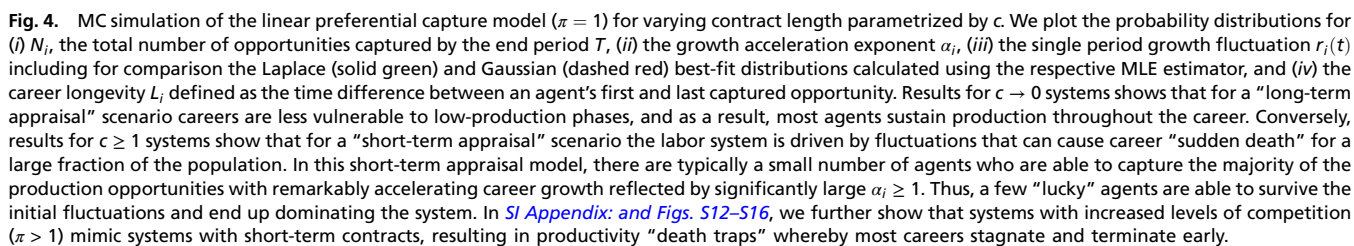
until all P opportunities for a given period t are captured. We assume that each agent has the production potential of one unit per period, and so the total number of opportunities distributed per period P is equal to the number of competing agents, $P \equiv I$.

We use Monte Carlo (MC) simulation to analyze this two-parameter model over the course of $t = 1 \dots T$ sequential periods. In each production period (i.e., representing a characteristic time to publication), a fixed number of P production units are captured by the competing agents. At the end of each period, we update each $w_i(t)$ and then proceed to simulate the next preferential capture period $t + 1$. Because $\mathcal{P}_i(t)$ depends on the relative achievements of every agent, the relative competitive advantage of one individual over another is determined by the parameter π . In the *SI Appendix* we elaborate in more detail the results of our simulation of synthetic careers dynamics. We vary π and c for a labor force of size $I \equiv 1,000$ and maximum lifetime $T \equiv 100$ periods as a representative size and duration of a real labor cohort. Our results are general, and for sufficiently large system size, the qualitative features of the results do not depend significantly on the choice of I or T .

The case with $\pi = 0$ corresponds to a random capture model that has (i) no appraisal and (ii) no preferential capture. Hence, in this null model, opportunities are captured at a Poisson rate $\lambda_p = 1$ per period. The results of this model (see *SI Appendix: Fig. S13*) show that almost all careers obtain the maximum career length T with a typical career trajectory exponent $\langle \alpha_i \rangle \approx 1$. Comparing to simulations with $\pi > 0$ and $c \geq 0$, the null model is similar to a “long-term” appraisal system ($c \rightarrow 0$) with sublinear preferential capture ($\pi < 1$). In such systems, the long-term appraisal time scale averages out fluctuations, and so careers are significantly less vulnerable to periods of low production and hence more sustainable because they are not determined primarily by early career fluctuations.

However, as π increases, the strength of competitive advantage in the system increases, and so some careers are “squeezed out” by the larger more dominant careers. This effect is compounded by short-term appraisal corresponding to $c \approx 1$. In such systems with superlinear capture rates and/or relatively large c , most individuals experience “sudden death” termination relatively early in the career. Meanwhile, a small number of “stars” survive the initial selection process, which is governed primarily by random chance, and dominate the system.

We found drastically different lifetime distributions when we varied the appraisal (contract) length (see *SI Appendix: Figs. S12–S16*). In the case of linear preferential capture with a long-term appraisal system $c = 0$, we find that 10% of the labor population terminates before reaching career age $0.94T$ (where T is the maximum career length or “retirement age”), and only 25% of the labor population terminates before reaching career age $0.98T$. On the contrary, in a short-term appraisal system with $c = 1$, we find that 10% of the labor population terminates



In an attempt to render a more objective review process for tenure and other lifetime achievement awards, quantitative measures for scientific publication impact are increasing in use and variety (17–20, 24, 27, 46, 47). However, many quantifiable benchmarks such as the *h*-index (17) do not take into account collaboration size or discipline specific factors. Measures for the comparison of scientific achievement should at least account

for variable collaboration, publication, and citation factors (19, 46, 47). Hence, such open problems call for further research into the quantitative aspects of scientific output using comprehensive longitudinal data for not just the extremely prolific scientists, but the entire labor force.

Current scientific trends indicate that there will be further increases in typical team sizes that will forward the emergent complexity arising from group dynamics (7, 12, 42), and overall, an incredible growth of science. There is an increasing need for individual/group production measures, such as the output measure Q , following from Eq. 7, which accounts for group efficiency factors. Normalized production measures which account for co-authorship factors have been proposed in refs. 19, 46, but the measures proposed therein do not account for the variations in team productivity.

The complexity of large collaborations raises open questions concerning scientific productivity and the organization of teams. We measure a decreasing marginal return $\gamma < 1$ with increasing group size which identifies the importance of team management. A theory of labor productivity can help improve our understanding of institutional growth, for organizations ranging in size from scientific collaborations to universities, firms, and countries (33, 34, 44, 47–50).

ACKNOWLEDGMENTS. We thank D. Helbing, N. Dimitri, O. Penner, and an anonymous PNAS Board Member for insightful comments. We gratefully acknowledge support from the IMT and Keck Foundations, the Defense Threat Reduction Agency (DTRA) and Office of Naval Research (ONR), and the National Science Foundation (NSF) Chemistry Division (Grants CHE 0911389 and CHE 0908218).

- David PA (2008) The Historical Origins of 'Open Science': an essay on patronage, reputation and common agency contracting in the scientific revolution. *Capitalism and Society* 3:1–103 Article 5.
- Helbing D, Ballelli S (2011) How to create an innovation accelerator. *Eur Phys J Spec Top* 195:101–136.
- Chait RP, ed. (2002) *The Questions of Tenure* (Harvard University Press, Cambridge, USA).
- Austin J (2011) Two new studies address jobs in STEM. *Science Career Magazine*, doi: 10.1126/science.caredit.a1100130.
- Cyranoski D, Gilbert N, Ledford H, Nayar A, Yahia M (2011) The PhD factory. *Nature* 472:276–279.
- Kaplan K (2010) The changing face of tenure. *Nature* 468:123–125.
- Wutchy S, Jones BF, Uzzi B (2008) The increasing dominance of teams in production of knowledge. *Science* 322:1036–1039.
- Petersen AM, Jung W-S, Stanley HE (2008) On the distribution of career longevity and the evolution of home-run prowess in professional baseball. *Europhys Lett* 83:50010, doi: 10.1209/0295-5075/83/50010.
- Petersen AM, Penner O, Stanley HE (2011) Methods for detrending success metrics to account for inflationary and deflationary factors. *Eur Phys J B* 79:67–78.
- Coates JM, Herbert J (2008) Endogenous steroids and financial risk taking on a London trading floor. *Proc Natl Acad Sci USA* 105:6167–6172.
- Saavedra S, Hagerty K, Uzzi B (2011) Synchronicity, instant messaging, and performance among financial traders. *Proc Natl Acad Sci USA* 108:5296–5301.
- Börner K, et al. (2010) A multi-level systems perspective for the science of team science. *Sci Transl Med* 2:49cm24.
- Ratkiewicz J, Fortunato S, Flammini A, Menczer F, Vespignani A (2010) Characterizing and modeling the dynamics of online popularity. *Phys Rev Lett* 105:158701.
- Shockley W (1957) On the statistics of individual variations of productivity in research laboratories. *Proceedings of the Institute of Radio Engineers* 45:279–190, doi: 10.1109/JRPROC.1957.278364.
- Simonton DK (1997) Creative productivity: a predictive and explanatory model of career trajectories and landmarks. *Psychol Rev* 104:66–89.
- Petersen AM, Jung W-S, Yang J-S, Stanley HE (2011) Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proc Natl Acad Sci USA* 108:18–23.
- Petersen AM, Stanley HE, Succi S (2011) Statistical regularities in the rank-citation profile of scientists. *Sci Rep* 1:181, doi: 10.1038/srep00181.
- Radicchi F, Fortunato S, Markines B, Vespignani A (2009) Diffusion of scientific credits and the ranking of scientists. *Phys Rev E* 80:056103.
- Petersen AM, Wang F, Stanley HE (2010) Methods for measuring the citations and productivity of scientists across time and discipline. *Phys Rev E* 81:036114.
- Benjamin FJ, Weinberg BA (2011) Age dynamics in scientific creativity. *Proc Natl Acad Sci USA* 108:18910–18914.
- Kaminski D, Geisler C (2012) Survival analysis of faculty retention in science and engineering by gender. *Science* 335:864–866.
- Merton RK (1968) The Matthew effect in science. *Science* 159:56–63.
- De Solla Price D (1976) A general theory of bibliometric and other cumulative advantage processes. *J Am Soc Inform Sci* 27:292–306.
- Wu J, Lozano S, Helbing D (2011) Empirical study of the growth dynamics in real career h-index sequences. *J Informetr* 5:489–497.
- Allison PD, Steward JA (1974) Productivity differences among scientists: evidence for accumulative advantage. *Amer Sociol Rev* 39:596–606.
- Vespignani A (2009) Predicting the behavior of techno-social systems. *Science* 325:425–428.
- Mazloumian A, Eom Y-H, Helbing D, Lozano S, Fortunato S (2011) How citation boosts promote scientific paradigm shifts and Nobel prizes. *PLoS ONE* 6:e18975.
- Azoulay P, Zivin JSG, Wang J (2010) Superstar extinction. *Q J Econ* 125:549–589.
- Long JS, Allison PD, McGinnis R (1979) Entrance into the academic career. *Amer Sociol Rev* 44:816–830.
- Cole S, Cole JR (1967) Scientific output and recognition: a study in the operation of the reward system in science. *Amer Sociol Rev* 32:377–390.
- Barabási AL (2005) The origin of bursts and heavy tails in human dynamics. *Nature* 435:207–211.
- Gabrielli A, Caldarelli G (2007) Invasion percolation and critical transient in the Barabási model of human dynamics. *Phys Rev Lett* 98:208701.
- Plerou V, Amaral LAN, Gopikrishnan P, Meyer M, Stanley HE (1999) Similarities between the growth dynamics of university research and of competitive economic activities. *Nature* 400:433–437.
- Fu D, et al. (2005) The growth of business firms: theoretical framework and empirical evidence. *Proc Natl Acad Sci USA* 102:18801–18806.
- Capocci A, Rao F, Caldarelli G (2008) Taxonomy and clustering in collaborative systems: the case of the on-line encyclopedia Wikipedia. *EPL* 81:28006, doi: 10.1209/0295-5075/81/28006.
- Romer PM (1987) Growth based on increasing returns due to specialization. *Am Econ Rev* 77:56–62.
- Owen-Smith J, Powell WW (2004) Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community. *Organ Sci* 15:5–21.
- Börner K, Maru JT, Goldstone RL (2004) The simultaneous evolution of author and paper networks. *Proc Natl Acad Sci USA* 101:5266–5273.
- Palla G, Barabási AL, Visek T (2007) Quantifying social group evolution. *Nature* 446:664–667.
- Newman MEJ (2004) Coauthorship network and patterns of scientific collaboration. *Proc Natl Acad Sci USA* 101:5200–5205.
- Catanzaro M, Caldarelli G, Pietronero L (2004) Assortative model for social networks. *Phys Rev E* 70:037101.
- Guimerá R, Uzzi B, Spiro J, Amaral LAN (2005) Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308:697–702.
- Yamasaki K, et al. (2006) Preferential attachment and growth dynamics in complex systems. *Phys Rev E* 74:035103.
- Buldyrev SV, Gnowiec J, Pammolli F, Riccaboni M, Stanley HE (2007) The growth of business firms: facts and theory. *J Eur Econ Assoc* 5:574–584.
- Arrow KJ (1962) Economic welfare and the allocation of resources for invention. *The Rate and Direction of Inventive Activity*, ed RR Nelson (Princeton University Press, Princeton, NJ), pp 609–625.
- Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: toward an objective measure of scientific impact. *Proc Natl Acad Sci USA* 105:17268–17272.
- Matia K, Amaral LAN, Luwel M, Moed HF, Stanley HE (2005) Scaling phenomena in the growth dynamics of scientific output. *Journal of the American Society for Information Science* 56:893–902.
- Stanley MHR, et al. (1996) Scaling behavior in the growth of companies. *Nature* 379:804–806.
- Riccaboni M, Pammolli F, Buldyrev SV, Ponta L, Stanley HE (2008) The size variance relationship of business firm growth rates. *Proc Natl Acad Sci USA* 105:19595–19600.
- Podobnik B, Horvatic D, Petersen AM, Njavro M, Stanley HE (2009) Quantitative relations between risk, return and firm size. *EPL* 85:50003, doi: 10.1209/0295-5075/85/50003.